END
DATE
FILMED
9 83
DTIC

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

AD A131906

DTIC FILE COPY

②

ASYMPTOTIC MAXIMAL DEVIATION OF M-SMOOTHERS

DTIC
ELECTE
AUG 29 1983
S   D
                B

by

Wolfgang Härdle

Universität Heidelberg   University of North Carolina
Sonderforschungsbereich 123   Department of Statistics
Jm Neuenheimer Feld 293   321 Phillips Hall 039A
D-6900 Heidelberg 1   Chapel Hill, NC 27514

April 1983

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

83 08 19 07 7

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 83-0688	2. GOVT ACCESSION NO. AD-A131 90b	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) "ASYMPTOTIC MAXIMAL DEVIATION OF M-SMOOTHERS"		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Wolfgang Härdle		8. CONTRACT OR GRANT NUMBER(s) F49620-82-C-0009
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of North Carolina Department of Statistics Chapel Hill, NC 27514		10. PROGRAM ELEMENT. PROJECT, TASK AREA & WORK UNIT NUMBERS PE61102F: 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR/NM Building 410 Bolling AFB DC 20332		12. REPORT DATE April 1983
		13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robust nonparametric regression, Kernel estimators, Uniform consistency, Robust smoothing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be iid rv's with pdf $f(x,y)$ and let $m(x) = E(Y|X = x) = \int yf(x,y)dy/f_X(x)$ be the regression function of Y on X. The function $m(x)$ is estimated by $m_n(x)$ a solution of $(nh)^{-1} \sum_{i=1}^{n} K((x-X_i)/h) \cdot (Y_i - \cdot) = C$ for some odd and bounded Ψ-function making $m_n(x)$ a robust estimate of $m(x)$. Probabilities of maximal deviation of $|m_n(x) - m(x)|$ are computed in a similar way as in Bickel and Rosenblatt (1973) for density estimation and in Johnston (1982) for nonparametric regression function estimation.

ASYMPTOTIC MAXIMAL DEVIATION OF M-SMOOTHERS*

Wolfgang Härdle

Universität Heidelberg
Sonderforschungsbereich 123
Jm Neuenheimer Feld 293
D-6900 Heidelberg 1

University of North Carolina
Department of Statistics
321 Phillips Hall 039A
Chapel Hill, NC 27514

ABSTRACT

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be iid rv's with pdf $f(x,y)$ and let $m(x) = E(Y|X = x) = \int y f(x,y) dy / f_X(x)$ be the regression function of Y on X. The function $m(x)$ is estimated by $m_n(x)$ a solution of $(nh)^{-1} \sum_{i=1}^{n} K((x-X_i)/h) \Psi(Y_i - \cdot) = 0$ for some odd and bounded Ψ-function making $m_n(x)$ a robust estimate of $m(x)$. Probabilities of maximal deviation of $|m_n(x) - m(x)|$ are computed in a similar way as in Bickel and Rosenblatt (1973) for density estimation and in Johnston (1982) for nonparametric regression function estimation.

1. BACKGROUND AND INTRODUCTION

Nadaraya (1964) and Watson (1964) independently proposed the following kernel estimator

(1.1) $$m_n^*(x) = (nh_n)^{-1} \sum_{i=1}^{n} K((x-X_i)/h_n)Y_i / [(nh_n)^{-1} \sum_{j=1}^{n} K((x-X_j)/h_n)]$$

of the regression function $m(x) = \int y\, f(x,y)dy / f_X(x)$ where $f_X(x)$ denotes the marginal density of $X, K(\cdot)$ is a kernel and $\{h_n\}$ is a sequence of positive constants ("bandwidth"). Basically this estimator averages the Y's around $X = x$ motivated from the integral formula for $m(x)$ above. The numerator is a weighted local average of the Y's while the denominator is a density esti-mate of $f_X(x)$.

It is clear that occasional outliers generated by heavy tailed condi-tional densities $f(y|x)$ introduce smooth peaks and troughs in the estimated curve $m_n^*(x)$. Such outliers occur quite often in practice. (Ruppert et al., 1982 Figure 7 or Bussian et al., 1982). To avoid this misleading property of $m_n^*(x)$ due to spiky Y-observations we introduce a robust estimate, the M-smoother, $m_n(x)$ as the solution of

(1.2) $$(nh_n)^{-1} \sum_{i=1}^{n} K((x-X_i)/h_n)\psi(Y_i - \cdot) = 0,$$

where ψ denotes a bounded, odd and continuous function. Note that if $\psi(u) = u$, then m_n is the Nadaraya-Watson estimator m_n^*. Bias and variance rates for $m_n(x)$ with K as the uniform window where obtained by Stuetzle and Mittal (1979), robustness properties, consistency and asymptotic normality of $m_n(x)$ were considered by Härdle (1982). For the case of nonrandom design, i.e. X_i attains fixed values, we may refer to Härdle and Gasser (1982). In this paper we show that

(1.3)
$$P\{(2\alpha \log n)^{\frac{1}{2}}[\sup_{0\leq t\leq 1} |(m_n(t) - m(t))\cdot r(t)|/\lambda(K)^{\frac{1}{2}} - d_n] < x\}$$

$$\cdot \exp(-2 \exp(-x)) \quad ,$$

where α, $r(t)$, $\lambda(K)$, d_n are suitable scaling parameters.

The result (1.3) improves upon that of Johnston (1982) in a number of ways. First, Johnston obtains results like (1.3), but for estimates different from the Nadaraya-Watson estimator (1.1); our result (1.3) of course applies to the Nadaraya-Watson estimator as a special case. Secondly, (1.3) holds for a much broader class of estimators. Finally, we obtain (1.3) under assumptions weaker than those needed by Johnston.

2. ASSUMPTIONS AND RESULTS

We write h for the bandwidth h_n from here on unless there is no need to do so. We make use of the following assumptions.

(A1) the kernel $K(\cdot)$ is positive has compact support $[-A,A]$ and is continuously differentiable.

(A2) $(nh)^{-\frac{1}{2}}(\log n)^{3/2} \to 0 \qquad (n \log n)^{\frac{1}{2}} h^{5/2} \to 0$

$(nh^3)^{-1}(\log n)^2 \leq M$, M a constant .

(A3) $h^{-3}(\log n) \int_{|y| > a_n} f_y(y)dy = O(1)$, $f_y(y)$ the marginal density

of Y, $\{a_n\}_{n=1}^{\infty}$ a sequence of constants tending to infinity as $n \to \infty$.

(A4) $\inf_{0 < t \leq 1} |q(t)| \geq q_0 > 0$, where $q(t) = E(\Psi'(Y-m(t))|X=t) - f_X(t)$

(A5) the regression function m(x) is twice continuously differentiable, the conditional densities $f(y|x)$ are symmetric for all x, Ψ is piecewise twice continuously differentiable.

We need some more definitions before we discuss the assumptions.

Define

$$\sigma^2(t) = E(\Psi^2(Y-m(t))|X=t)$$

$$H_n(t) = (nh)^{-1} \sum_{i=1}^{n} K((x-X_i)/h)\Psi(Y_i-m(t))$$

$$D_n(t) = (nh)^{-1} \sum_{i=1}^{n} K((x-X_i)/h)\Psi'(Y_i-m(t)).$$

We further assume that $\sigma^2(t)$ and $f_X(t)$ are differentiable.

Assumption (A1) on the compact support of the kernel could possibly be relaxed introducing a cutoff technique as Csörgö and Hall (1982) for density estimators. Assumption (A2) has purely technical reasons: to keep the bias down and to ensure the vanishing of the nonlinear remainder terms. Assumption (A3) appears in a somewhat modified form also in Johnston's paper (1982). When we want to apply the following theorem to the Nadaraya-Watson estimator $m_n^*(x)$ we have actually to restate (A2) as h^{-3} $(\log n)$ $\int_{|y|>a_n} y^2 f_y(y)dy$ (which is assumption A1 in Johnston (1982)). Assumption (A5) stating the symmetry of the conditional densities is common in robustness considerations (Huber, 1981). It guarantees that the only solution of $\int \Psi(y-\cdot)f(y|x)dy = 0$ is $m(x) = E(Y|X=x)$. If we had skew distributions then we would no longer estimate the conditional mean but rather a conditional quantile such as the median.

Theorem

Let $h = n^{-\delta}$, $1/5 < \delta < 1/3$ and $\lambda(K) = \int_{-A}^{A} K^2(u)du$ and

$$d_n = (2\delta \log n)^{\frac{1}{2}} + (2\delta \log n)^{-\frac{1}{2}}\{\log(c_1(K)/\Pi^{\frac{1}{2}}) + \frac{1}{2}[\log \delta + \log \log n]\},$$
$$\text{if } c_1(K) = K^2(A) + K^2(-A)/[2\lambda(K)] > 0$$

$$d_n = (2\delta \log n)^{\frac{1}{2}} + (2\delta \log n)^{-\frac{1}{2}} \{\log (c_2(K)/2\Pi)\}$$
$$\text{otherwise with } c_2(K) = \int_{-A}^{A}[K'(u)]^2 du/[2\lambda(K)].$$

Then (1.3) holds with

$$r(t) = (nh)^{\frac{1}{2}}q(t)[\sigma^2(t)f_X(t)]^{-\frac{1}{2}} .$$

This theorem can be used to construct uniform confidence intervals for the regression function as stated in the following corollary.

Corollary: Assuming the theorem above holds, an approximate $(1-\alpha) \times 100\%$ confidence band over $[0,1]$ is

$$m_n(t) \pm (nh)^{-\frac{1}{2}}[\sigma^2(t)f_X(t)\lambda(K)]^{\frac{1}{2}} q^{-1}(t)[d_n+c(\alpha)(2\delta \log n)^{-\frac{1}{2}}] \cdot [\lambda(K)]^{\frac{1}{2}}$$

where $c(\alpha) = \log 2 - \log|\log(1-\alpha)|$.

The proof is essentially based on a linearization argument due to Taylor series expansion. The leading linear term will then be approximated in a similar way as in Johnston (1982), Bickel and Rosenblatt (1973). The main idea behind the proof is a strong approximation of the empirical process of $\{(X_i,Y_i)\}_{i=1}^n$ by a sequence of Brownian bridges (with two dimensional time) as provided Tusnady (1977).

It follows by Taylor expansions applied to the defining equation (1.2) that

(2.1) $$m_n(t) - m(t) = (H_n(t)-EH_n(t))/q(t) + R_n(t)$$

where $[H_n(t)-EH_n(t)]/q(t)$ is the leading linear term and

(2.2) $$R_n(t) = H_n(t)[q(t)-D_n(t)]/[D_n(t)\cdot q(t)] + EH_n(t)/q(t)$$

$$+ \frac{1}{2}(m_n(t)-m(t))^2 \cdot [D_n(t)]^{-1}\cdot (nh)^{-1} \sum_{i=1}^{n} K((x-X_i)/h)\Psi''(Y_i-m(t)+r_n^{(i)}(t)),$$

$$|r_n^{(i)}(t)| \le |m_n(t)-m(t)|.$$

is the remainder term. In the third section it is shown (Lemma 3.1) that

$$\| R_n \| = \sup_{0 \le t \le 1} |R_n(t)| = o_p((nh \log n)^{-\frac{1}{2}}).$$

Furthermore the rescaled linear part

$$Y_n(t) = (nh)^{\frac{1}{2}}[\sigma^2(t)f_X(t)]^{-\frac{1}{2}}(H_n(t) - EH_n(t))$$

is approximated by a sequence of Gaussian processes, leading finally to the following process

$$Y_{5,n}(t) = h^{-\frac{1}{2}} \int K((t-x)/h) \, dW(x),$$

as in Bickel and Rosenblatt (1973).

We also need the Rosenblatt transformation (Rosenblatt, 1952).

$$T(x,y) = (F_{X|y}(x|y), \, F_Y(y))$$

which transforms (X_i, Y_i) into $T(X_i, Y_i) = (X_i', Y_i')$ mutually independent uniform rv's. With the aid of this transformation Theorem 1 of Tusnàdy (1977) may be applied to obtain the following lemma.

Lemma 2.1: On a suitable probability space there exists a sequence of Brownian bridges B_n such that

$$\sup_{x,y} |Z_n(x,y) - B_n(T(x,y))| = O(n^{-\frac{1}{2}}(\log n)^2) \quad \text{a.s.},$$

where $Z_n(x,y) = n^{\frac{1}{2}}[F_n(x,y) - F(x,y)]$ denotes the empirical process of $\{(X_i, Y_i)\}_{i=1}^{n}$.

Before we define the different approximating processes let us first rewrite $Y_n(t)$ as a stochastic integral with respect to the empirical process $Z_n(x,y)$.

$$Y_n(t) = h^{-\frac{1}{2}} g'(t)^{-\frac{1}{2}} \int\int K((t-x)/h) \Psi(y-m(t)) dZ_n(x,y), \quad g'(t) = \sigma^2(t) \, f_X(t).$$

The approximating processes are now

$$Y_{0,n}(t) = (hg(t))^{-\frac{1}{2}} \int\int_{\Gamma_n} K((t-x)/h) \Psi(y-m(t)) dZ_n(x,y),$$

where $\Gamma_n = \{|y| \le a_n\}$, $g(t) = E(\Psi^2(y-m(t)) \cdot I(|y| \le a_n) | X=t) \cdot f_X(t)$

$$Y_{1,n}(t) = (hg(t))^{-\frac{1}{2}} \int\int_{\Gamma_n} K((t-x)/h) \Psi(y-m(t)) \, dB_n(T(x,y)),$$

$\{B_n\}$ being the sequence of Brownian bridges from Lemma 2.1.

$$Y_{2,n}(t) = (hg(t))^{-\frac{1}{2}} \int\int_{\Gamma_n} K((t-x)/h) \Psi(y-m(t)) \, dW_n(T(x,y))$$

$\{W_n\}$ being the sequence of Wiener processes satisfying

$$B_n(x',y') = W_n(x', y') - x'y'W_n(1,1)$$

$$Y_{3,n}(t) = (hg(t))^{-\frac{1}{2}} \iint_n K((t-x)/h) \psi(y-m(x)) \, dW_n(T(x,y))$$

$$Y_{4,n}(t) = (hg(t))^{-\frac{1}{2}} \int g(x)^{\frac{1}{2}} K((t-x)/h) dW(x)$$

$$Y_{5,n}(t) = h^{-\frac{1}{2}} \int K((t-x)/h) dW(x),$$

$\{W(\cdot)\}$ being the Wiener process on $(-\infty, \cdot)$.

Lemmata 3.2 to 3.7 ensure that all these processes have the same limit distributions. The results then follows from the following lemma

Lemma 2.2 (Bickel and Rosenblatt (1973)). Let d_n, $\lambda(K)$, δ as in the theorem.

Let

$$Y_{5,n}(t) = h^{-\frac{1}{2}} \int K((t-x)/h) \, dW(x).$$

Then

$$P((2\delta \log n)^{\frac{1}{2}} \{ \sup_{0 \leq t \leq 1} | Y_{5,n}(t) | / [\lambda(K)]^{\frac{1}{2}} - d_n \} < x) \to e^{-2e^{-x}} \quad .$$

3. PROOFS

We show first that $\| R_n \| = \sup_{0 \leq t \leq 1} |R_n(t)|$ vanishes asymptotically with the desired rate $(nh \log n)^{-\frac{1}{2}}$.

Lemma 3.1: For the remainder term $R_n(t)$ defined in (2.2) we have

$$(3.1) \qquad \| R_n \| = o_p((nh \log n)^{-\frac{1}{2}}) \quad .$$

Proof: First we have by the positivity of the kernel K and $|\psi'''| < C_1$

$$\| R_n \| \leq [\inf_{0 \leq t \leq 1} (|D_n(t)| \cdot q(t))]^{-1} \{ \| H_n \| \cdot \| q - D_n \| + \| D_n \| \cdot \| EH_n \| \}$$

$$+ C_1 \cdot \| m_n - m \|^2 \cdot [\inf_{0 \leq t \leq 1} |D_n(t)|]^{-1} \cdot \| f_n \| ,$$

where $f_n = (nh)^{-1} \sum_{i=1}^{n} K((x-X_i)/h)$.

The desired result (3.1) will then follow if we prove the following:

$$(3.2) \qquad \| H_n \| = o_p(n^{-\frac{1}{4}} h^{-\frac{1}{4}} \cdot (\log n)^{-\frac{1}{2}}) \tag{3.2}$$

(3.3) $\quad \|q-D_n\| = o_p(n^{-\frac{1}{4}}h^{-\frac{1}{4}}(\log n)^{-\frac{1}{2}})$

(3.4) $\quad \|FH_n\| = o(h^2)$

(3.5) $\quad \|m_n-m\|^2 = o_p((nh)^{-\frac{1}{2}}(\log n)^{-\frac{1}{2}}).$

Define $U_n(t) = n^{\frac{1}{4}}h^{\frac{1}{4}}(\log n)^{\frac{1}{2}}[H_n(t)-EH_n(t)].$

We first show that $U_n(t) \overset{P}{\to} 0$ for all t. This follows from Markov's inequality since

$$U_n(t) = \sum_{i=1}^{n} U_{i,n}(t) ,$$

where $U_{i,n}(t) = n^{-3/4}h^{-3/4}(\log n)^{\frac{1}{2}}[K((t-X_i)/h)\psi(Y_i-m(t))-EK((t-X)/h)\cdot\psi(y-m(t))],$

are iid rv's and thus

$$P(|U_n(t)|>\epsilon) < \epsilon^{-2}n^{-\frac{1}{2}}h^{-\frac{1}{2}}(\log n)\cdot h^{-1}EK^2((t-X)/h)\psi^2(Y-m(t)).$$

The RHS of this inequality tends to zero since

$$h^{-1}EK^2((t-X)/h)\psi^2(Y-m(t)) = h^{-1}\int K^2((t-u)/h)E(\psi^2(Y-m(t))|X=u)f_X(u)du$$
$$\sim \sigma^2(t)\cdot f_X(t)\cdot\int K^2(u)du$$

by continuity of $\sigma^2(t)$ and $f_X(t)$.

Next we show the tightness of $U_n(t)$ using the following moment condition (Billingsley, 1968, Th. 15.6)

$$E\{|U_n(t)-U_n(t_1)|\cdot|U_n(t_2)-U_n(t)|\} \leq C_2\cdot(t_2-t_1)^2$$

where C_2 is a constant.

By the Schwarz inequality,

$$E\{|U_n(t)-U_n(t_1)|\cdot|U_n(t_2)-U_n(t)|\}$$
$$\leq \{E[U_n(t)-U_n(t_1)]^2\cdot E[U_n(t_2)-U_n(t)]^2\}^{\frac{1}{2}} .$$

It suffices to consider only the term $E\{U_n(t)-U_n(t_1)\}^2$.

Using the Lipschitz continuity of K,ψ,m and assumption (A2) we have

$$\{E[U_n(t)-U_n(t_1)]^2\}^{\frac{1}{2}}$$

$$\cdot (\log n)(nh)^{-3/2} \cdot E[A+B]^{2 \cdot \frac{1}{2}}$$

$$\cdot C_A(nh)^{-\frac{1}{2}}(\log n)^{\frac{1}{2}} \, t-t_1 \, + C_B(n^{-\frac{1}{2}}h^{-3/4}(\log n)^{\frac{1}{2}} \cdot \, t-t_1 \, \cdot C_3 \cdot \, t-t_1$$

where $A = \sum_{i=1}^{n} K((t-X_i)/h)[\psi(Y_i-m(t))-\psi(Y_i-m(t_1))]$

$B = \sum_{i=1}^{n} \cdot (Y_i-m(t_1))[K((t_1-X_i)/h)-K((t-X_i)/h)],$

and C_A, C_B are Lipschitz bounds for ψ, m, K.

Since (3.4) follows from the well-known bias calculation

$$EH_n(t) = h^{-1} \int K((t-u)/h)E(\psi(y-m(t)) \, X=u)f_X(u)du = O(h^2),$$

where $O(h^2)$ is independent of t (Parzen, 1962) we have from assumption (A2)

that $\, EH_n \, = o((nh)^{-\frac{1}{2}}(\log n)^{-\frac{1}{2}})$.

Statement (3.2) thus follows using tightness of $U_n(t)$ and the inequality

$\, H_n \, \cdot \, H_n-EH_n \, + \, EH_n \, $.

Statement (3.3) follows in the same way as (3.2) using assumption (A2)

and the continuity properties of K, ψ', m.

Finally from Härdle and Luckhaus (1982), where uniform continuity of

$m_n(t)-m(t)$ is shown, we have

$$\, m_n-m \, = O_p((nh)^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}),$$

which implies (3.5) .

Now the assertion of the lemma follows since by tightness of $D_n(t)$,

$\inf_{0 \cdot t \cdot 1} \, D_n(t) \, \frac{1}{p} \cdot q_0$ and thus

$$\| R_n \| = o_p((nh)^{-\frac{1}{2}}(\log n)^{-\frac{1}{2}})(1 + \| f_n \|).$$

Finally by Theorem 3.1 of Bickel and Rosenblatt (1973) $\| f_n \| = O_p(1)$,

thus the desired result $\| R_n \| = o_p((nh)^{-\frac{1}{2}}(\log n)^{-\frac{1}{2}})$ follows. In the nonrobust

case, i.e. $\psi(u) = u$, the remainder term R_n reads

(3.6) $\qquad R_n = [m_n^\star - m][f_X-f_n]f_X^{-1} + E(\hat{m}_n-mf_n)/f_X$,

where $\dot{m}_n(x) = (nh)^{-1} \sum_{i=1}^{n} K((x-X_i)/h)Y_i$.

Johnston (1982) proved that $(\hat{m}_n - E\,\hat{m}_n)/f$ has the desired asymptotic distribution as stated in our Theorem.

So if we apply the recent result of Mack and Silverman (1982) or Härdle and Luckhaus (1982) to $\| m_n^* - m \|$ and the well known result from Bickel and Rosenblatt (1973) to $\| f_X - f_n \|$ we may conclude that the first term on the RHS of (3.6) is $o_p((nh)^{-\frac{1}{2}}(\log n)^{-\frac{1}{2}})$. The second term in (3.6) is

$$[h^{-1}\int K((t-u)/h)\cdot m(u)f(u)du - m(t)h^{-1}\int K((t-u)/h)f(u)du]/f_X(t)$$

which is by the same calculations as mentioned above (Parzen, 1962) of the order $O(h^2)$. This shows that our result generalizes Johnston's paper. Our theorem says also that the confidence bounds are smaller. Johnston had $s^2(t) = E(Y^2|X=t)$ as a factor for the asymptotic confidence bound, we have $\sigma^2(t) = \text{var}(Y|X=t)$ which is in general smaller than $s^2(t)$. We now begin with the subsequent approximations of the processes $Y_{0,n}$ to $Y_{5,n}$.

Lemma 3.2:
$$\| Y_{0,n} - Y_{1,n} \| = O((nh)^{-\frac{1}{2}}(\log n)^2) \qquad\qquad \text{a.s.}$$

Proof: Let t be fixed and put $L(y) = \Psi(y - m(t))$ still depending on t.

Use integration by parts and obtain:
$$\int_{\Gamma_n}\int L(y)K((t-x)/h)dZ_n(x,y) =$$

$$= \int_{u=-A}^{A}\int_{y=-a_n}^{a_n} L(y)K(u)dZ_n(t-h\cdot u,y) =$$

$$= \int_{-A}^{A}\int_{-a_n}^{a_n} Z_n(t-h\cdot u,y)d[L(y)K(u)] + L(a_n)\int_{-A}^{A} Z_n(t-h\cdot u,a_n)dK(u)$$

$$-L(-a_n)\int_{-A}^{A} Z_n(t-h\cdot u,-a_n)dK(u) + K(A)[\int_{-a_n}^{a_n} Z_n(t-h\cdot u,y)dL(y)$$

$$+L(a_n)Z_n(t-h\cdot A,a_n)-L(-a_n)Z_n(t-h\cdot A,-a_n)]$$

$$-K(-A)[\int_{-a_n}^{a_n} Z_n(t+h\cdot A,y)dL(y) + L(a_n)Z_n(t+h\cdot A,a_n)$$

$$-L(-a_n)Z_n(t+h\cdot A,-a_n)] .$$

If we apply the same operations to $Y_{1,n}$ with $B_n(T(x,y))$ instead of $Z_n(x,y)$ and use Lemma 2.1 we finally obtain

$$\sup_{0 \cdot t \cdot 1} h^{\frac{1}{2}} q(t)^{\frac{1}{2}} |Y_{0,n}^{(t)} - Y_{1,n}(t)| = O((nh)^{-\frac{1}{2}}(\log n)^2) \qquad \text{a.s.}$$

using the differentiability and boundedness of ψ.

Lemma 3.3:

$$\| Y_{1,n} - Y_{2,n} \| = O_p(h^{\frac{1}{2}})$$

Proof: Note that the Jacobi of $T(x,y)$ is $f(x,y)$ hence

$$|Y_{1,n}(t) - Y_{2,n}(t)| = |(q(t)h)^{-\frac{1}{2}} \iint_n \psi(y-m(t)K((t-x)/h)f(x,y)dxdy| \cdot |W_n(1,1)|$$

It follows that

$$h^{-\frac{1}{2}} \| Y_{1,n} - Y_{2,n} \| \cdot |W_n(1,1)| \cdot \| g^{-\frac{1}{2}} \| \cdot \sup_{0 \le t \le 1} h^{-1} \iint_n \psi(y-m(t))K((t-x)/h)|f(x,y)dxdy$$

Since $\| g^{-\frac{1}{2}} \|$ is bounded by assumption and ψ is bounded we have

$$h^{-\frac{1}{2}} \| Y_{1,n} - Y_{2,n} \| \le |W_n(1,1)| \cdot C_4 \cdot h^{-1} \int (K((t-x)/h))dx = O_p(1).$$

Lemma 3.4:

$$\| Y_{2,n} - Y_{3,n} \| = O_p(h^{\frac{1}{2}})$$

Proof: The difference $|Y_{2,n}(t) - Y_{3,n}(t)|$ may be written as

$$|(q(t)h)^{-\frac{1}{2}} \iint_n [\psi(y-m(t)) - \psi(y-m(x)]K((t-x)/h)dW_n(T(x,y))|$$

If we use the fact that ψ, m are uniformly continuous this is smaller than

$$h^{-\frac{1}{2}} |q(t)|^{-\frac{1}{2}} \cdot O_p(h)$$

and the lemma thus follows.

Lemma 3.5:

$$\| Y_{4,n} - Y_{5,n} \| = O_p(h^{\frac{1}{2}})$$

Proof:

$$|Y_{4,n}(t) - Y_{5,n}(t)| = h^{-\frac{1}{2}} |\int \{[\frac{q(x)}{g(t)}]^{\frac{1}{2}} - 1\}K((t-x)/h)dW(x)| \le$$

$$
\cdot\ h^{-\frac{1}{2}} |\int_{-A}^{A} W(t-hu)\ \cdot\ \{[\frac{g(t-hu)}{g(t)}]^{\frac{1}{2}} - 1\}\ K(u)du|
$$

$$
+\ h^{-\frac{1}{2}} |K(A)W(t-hA)\{[\frac{g(t-Ah)}{g(t)}]^{\frac{1}{2}} - 1\}|
$$

$$
+\ h^{-\frac{1}{2}} |K(-A)W(t+hA)\{[\frac{g(t+h)A)}{g(t)}]^{\frac{1}{2}} - 1\}|
$$

$$
=\ S_{1,n}(t) + S_{2,n}(t) + S_{3,n}(t)\ ,\ \text{say}.
$$

The second term can be estimated by

$$
h^{-\frac{1}{2}} \| S_{2,n} \| \le K(A) \cdot \sup_{0 \le t \le 1} |W(t-Ah)| \cdot \sup_{0 \le t \le 1} h^{-1}|\{[\frac{g(t-Ah)}{g(t)}]^{\frac{1}{2}} - 1\}|
$$

by the mean value theorem it follows that

$$
h^{-\frac{1}{2}} \| S_{2,n} \| = 0_p(1).
$$

The first term $S_{1,n}$ is estimated as follows.

$$
h^{-1}S_{1,n}(t) = |h^{-1}\int_{-A}^{A} W(t-uh)\{K'(u)([\frac{g(t-uh)}{g(t)}]^{\frac{1}{2}} - 1\}\ du
$$

$$
-\frac{1}{2}\int_{-A}^{A} W(t-uh)K(u)[\frac{g(t-uh)}{g(t)}]^{-\frac{1}{2}}[\frac{g'(t-uh)}{g(t)}]du|
$$

$$
=\ |T_{1,n}(t) - T_{2,n}(t)|\ ,\ \text{say}.
$$

$$
\| T_{2,n} \| \le C_5 \cdot \int_{-A}^{A} |W(t-hu)|du = 0_p(1) \text{ by assumption on } g(t) = \sigma^2(t) \cdot f_\chi(t).
$$

To estimate $T_{1,n}$ we again use the mean value theorem to conclude that

$$
\sup_{0 \le t \le 1} h^{-1}|[\frac{g(t-uh)}{g(t)}]^{\frac{1}{2}} - 1| < C_6 \cdot |u|
$$

hence

$$
\| T_{1,n} \| \le C_6 \cdot \sup_{0 \le t \le 1} \int_{-A}^{A} |W(t-hu)K'(u)u|du = 0_p(1).
$$

Since $S_{3,n}(t)$ is estimated as $S_{2,n}(t)$ we finally obtain the desired result.

The next lemma shows that the truncation introduced through $\{a_n\}$ does not affect the limiting distribution.

Lemma 3.6:

$$\| Y_n - Y_{0,n} \| = 0_p((\log n)^{-\frac{1}{2}}).$$

Proof: We shall only show that $g'(t)^{-\frac{1}{2}} h^{-\frac{1}{2}} \iint\limits_{\mathbb{R}-\Gamma_n} \psi(y-m(t))K((t-x)/h)dZ_n(x,y)$

fulfills the lemma.

The replacement of $q'(t)$ by $g(t)$ may be proved as in Johnston (1982). The

quantity above is less than $h^{-\frac{1}{2}} \| g^{-\frac{1}{2}} \| \cdot \| \iint\limits_{\{|y|>a_n\}} \psi(y-m(\cdot))K((\cdot-x)/h)dZ(x,y) \|$.

It remains to show that the last factor tends to zero at a rate $0_p((\log n)^{-\frac{1}{2}})$.

We show first that

$$V_n(t) = (\log n)^{\frac{1}{2}} h^{-\frac{1}{2}} \iint\limits_{\{|y|>a_n\}} \psi(y-m(t))K((t-x)/h)dZ_n(x,y)$$

$$\xrightarrow{P} 0 \quad \text{for all } t$$

and then we show tightness of $V_n(t)$, the result then follows.

$$V_n(t) = (\log n)^{\frac{1}{2}}(nh)^{-\frac{1}{2}} \sum_{i=1}^{n} \{\psi(Y_i-m(t)I_{\{|y|>a_n\}}(Y_i)K((t-X_i)/h)$$

$$- E\psi(Y_i-m(t)) \cdot I_{\{|y|>a_n\}}(Y_i)K((t-X_i)/h)\}$$

$$= \sum_{i=1}^{n} X_{n,i}(t)$$

where $\{X_{n,i}(t)\}_{i=1}^{n}$ are iid for each n with $EX_{n,i}(t) = 0$ for all $t \in [0,1]$.

We have then

$$EX_{n,i}^2(t) \le (\log n)(nh)^{-1} E\psi^2(Y_i-m(t))I_{\{|y|>a_n\}}(Y_i)K^2((t-X_i)/h)$$

$$\le \sup_{-A \le u \le A} K^2(u) \cdot (\log n)(nh)^{-1} E\psi^2(Y_i- m(t))I_{\{|y|>a_n\}}(Y_i)$$

hence

$$\text{var}\{V_n(t)\} = E(\sum_{i=1}^{n} X_{n,i}(t))^2 = n \cdot EX_{n,i}^2(t)$$

$$\le \sup_{-A \le u \le A} K^2(u)h^{-1}(\log n) \int\limits_{\{|y|>a_n\}} f_y(y)dy \cdot M_\psi$$

where M_ψ denotes an upper bound for ψ^2.

This term tends to zero by assumption (A3). Thus by Markov's inequality we conclude that

$$V_n(t) \xrightarrow{\ p\ } 0 \qquad \text{for all } t \in [0,1].$$

To prove tightness of $\{V_n(t)\}$ we refer again to the following moment condition as stated in Lemma 3.1.

$$E\{|V_n(t) - V_n(t_1)| \cdot |V_n(t_2) - V_n(t)|\} \le C' \cdot (t_2 - t_1)^2$$

C' denoting a constant, $t \in [t_1, t_2]$.

We again estimate the left hand side by Schwarz's inequality and estimate each factor separately.

$$E[V_n(t) - V_n(t_1)]^2 = (\log n)(nh)^{-1} E\{ \sum_{i=1}^{n} \Psi_n(t, t_1, X_i, Y_i) \cdot I_{\{|y| > a_n\}}(Y_i) -$$
$$- E(\Psi_n(t, t_1, X_i, Y_i) \cdot I_{\{|y| > a_n\}}(Y_i))\}^2 ,$$

where $\Psi_n(t, t_1, X_i, Y_i) = \psi(Y_i - m(t)) K((t - X_i)/h) - \psi(Y_i - m(t_1)) K((t_1 - X_1)/h)$

Since ψ, m, K are Lipschitz continuous it follows

$$\{E[V_n(t) - V_n(t_1)]^2\}^{\frac{1}{2}}$$
$$\le C_7 \cdot (\log n)^{\frac{1}{2}} h^{-3/2} |t - t_1| \cdot \{ \int_{\{|y| > a_n\}} f_y(y) dy \}^{\frac{1}{2}}$$

If we apply the same estimations to $V_n(t_2) - V_n(t_1)$ we finally have

$$E\{|V_n(t) - V_n(t_1)| \cdot |V_n(t_2) - V_n(t)|\} \le C_7^2 (\log n) h^{-3} |t - t_1| | t_2 - t|$$
$$\cdot \int_{\{|y| > a_n\}} f_y(y) dy$$
$$\le C' \cdot |t_2 - t_1|^2 \text{ since } t \in [t_1, t_2] .$$
$$\text{by assumption (A3).}$$

Lemma 3.7: Let $\lambda(K) = \int K^2(u) du$ and let $\{d_n\}$ as in the theorem. Then

$$(2\delta \log n)^{\frac{1}{2}} [\| Y_{3,n} \| / [\lambda(K)]^{\frac{1}{2}} - d_n]$$

has the same asymptotic distribution as

$$(2\delta \log n)^{\frac{1}{2}} [\| Y_{4,n} \| / [\lambda(K)]^{\frac{1}{2}} - d_n]$$

Proof: $Y_{3,n}(t)$ is a Gaussian process with

$$EY_{3,n}(t) = 0$$

and covariance function

$$r_3(t_1,t_2) = EY_{3,n}(t_1)Y_{3,n}(t_2)$$

$$= [g(t_1)g(t_2)]^{-\frac{1}{2}}h^{-1}\iint_{\Gamma_n}\psi^2(y-m(x))K((t_1-x)/h)K((t_2-x)/h)f(x,y)dxdy.$$

$$= h^{-1}[g(t_1)g(t_2)]^{-\frac{1}{2}}\iint_{\Gamma_n}\psi^2(y-m(x))f(y|x)dyK((t_1-x)/h)K((t_2-x)/h)f_X(x)dx$$

$$= h^{-1}[g(t_1)g(t_2)]^{-\frac{1}{2}}\int g(x)K((t_1-x)/h)K((t_2-x)/h)\ dx$$

$$= r_4(t_1,t_2)$$ the covariance function of the Gaussian process $Y_{4,n}(t)$, which proves the lemma.

ACKNOWLEDGEMENT

REFERENCES

[1] Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviation of density function estimators. Ann. Stat. 1, 1071-1095.

[2] Billingsley, P. (1968). Convergence of Probability Measures. John Wiley & Sons.

[3] Bussian, B.M. and Härdle, W. (1982). Robust smoothing applied to White noise and single outlier contaminated Raman Spectra. Tentatively accepted by Applied Spectroscopy.

[4] Csörgö, S. and Hall, P. (1982). Upper and Lower Classes for Triangular Arrays. Z. F. Wahrscheinlichkeitstheorie und verw. Gebiete, 61, 207-222.

[5] Härdle, W. (1982). Robust regression function estimation. (To appear in J. Mult. Analysis).

[6] Härdle, W. and Gasser, T. (1982). Robust nonparametric function fitting. Submitted to JRSS(B).

[7] Härdle, W. and Luckhaus, S. (1982). Uniform consistency of a class of regression function estimates. Preprint SFB 123 University of Heidelberg #143.

[8] Huber, P. (1981). Robust statistics. J Wiley & Sons, New York.

[9] Johnston, G. (1982). Probabilities of maximal deviation of nonparametric regression function estimation. J. Mult. Anal. 12, 402-414.

[10] Mack, Y.P. and Silverman, B.W. (1982). Weak and strong uniform consistency of kernel regression estimates. Z. F. Wahrscheinlichkeitstheorie und verw. Gebiete, 61, 405-415.

[11] Nadaraya, E.A. (1964). On estimating regression. Theor. Prob. Appl. 9, 141-142.

[12] Parzen, M. (1962). On estimation of a probability density function and mode. Ann. Math. Stat. 32, 1065-1076.

[13] Rosenblatt, M. (1952). Remarks on a multivariate transformation. Ann. Math. Stat. 23, 470-472.

[14] Rosenblatt, M. (1971). Curve estimates. Ann. Math. Stat. 42, 1815-1842.

[15] Ruppert, D., Reish, R.L., Deriso, R.B., Carroll, R.J. (1982). Monte Carlo optimization by stochastic approximation (with application to harvesting of Atlantic menhaden). Inst. of Statistics Mimeo Series #1500, University of North Carolina at Chapel Hill.

[16] Stuetzle, W. and Mittal, Y. (1979). Some comments on the asymptotic behaviour of robust smoothers in "Smoothing Techniques for Curve Estimation" ed. T. Gasser and M. Rosenblatt. Springer Lecture Notes 757, Heidelberg.

[16] Tusnady, G. (1977). A remark on the approximation of the sample distri-
      bution function in the multidimensional case. <u>Period. Math. Hung</u>., 8,
      53-55.

[17] Watson, G.S. (1964). Smooth regression analysis. <u>Sankhya, Series</u> A,
      Vol. 26, 359-372.